PB90-232331

Statistical Issues in Epidemiologic Studies of the Health Effects of Ambient Acid Aerosols

Harvard School of Public Health, Boston, MA

Prepared for:

Health Effects Research Lab., Research Triangle Park, NC

1988

U.S. Department of Commerce
National Technical Information Service
NTIS

MS-90-049

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before complet

1. REPORT NO.	2.	3.
EPA/600/J-88/525		PB90-232331

4. TITLE AND SUBTITLE	5. REPORT DATE
Statistical Issues in Epidemiologic Studies of the Health Effects of Ambient Acid Aerosols.	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) James H. Ware	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Biostatistics Department Harvard School of Public Health Boston, Massachusetts	10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO. CR813495-01
12. SPONSORING AGENCY NAME AND ADDRESS Health Effects Research Laboratory - RTP, NC Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, NC 27711	13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE EPA/600/10

15. SUPPLEMENTARY NOTES Project Officer - John Creason
The Canadian Journal of Statistics, Vol 16(1), pp. 5-13; 3/88

16. ABSTRACT

Although the ecologic effects of acid rain have been widely reported, relatively little is known about the effects of acidic air pollution on human health. Some epidemiologic and animal studies suggest, however, that acidity is an important determinant of the respiratory health effects of aerosols. This paper reviews some of that evidence and discusses its implications for the design and analysis of epidemiologic studies. Two types of exposure patterns are contrasted: peak exposures associated with air pollution episodes, and chronic exposures resulting form persistently high levels of air pollutants. Recent work on the analysis of repeated categorical outcome variables provides new methods for the analysis of episode studies. Studies of long-term exposure require comparisons among population groups, and these comparisons can be subject to the design effects characteristic of multistage sample surveys. This paper examines the implications of these design effects for epidemiologic studies. The paper also discusses the measurement errors induced by the use of outdoor measurements to quantify personal exposure, and ways to assess effects of such errors on conventional analyses of air pollution studies.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT RELEASE TO THE PUBLIC	19. SECURITY CLASS (This Report) UNCLASSIFIED	21. NO. OF PAGES 10
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE A03

EPA Form 2220-1 (Rev. 4-77) PRI

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL INFORMATION SERVICE
SPRINGFIELD, VA. 22161

EPA/600/J-88/525

The Canadian Journal of Statistics
Vol. 16, No. 1, 1988, Pages 5–13
La Revue Canadienne de Statistique

Statistical issues in epidemiologic studies of the health effects of ambient acid aerosols*

James H. WARE and Daniel O. STRAM

Harvard School of Public Health

Key words and phrases: Ecological effects, air pollution, epidemiology, acidic aerosols, errors-in-variables, repeated measures.
AMS 1985 subject classifications: Primary 62P10, secondary 62K99.

ABSTRACT

Although the ecologic effects of acid rain have been widely reported, relatively little is known about the effects of acidic air pollution on human health. Some epidemiologic and animal studies suggest, however, that acidity is an important determinant of the respiratory health effects of aerosols. This paper reviews some of that evidence and discusses its implications for the design and analysis of epidemiologic studies. We contrast two types of exposure patterns: peak exposures associated with air pollution episodes, and chronic exposures resulting from persistently high levels of air pollutants. Recent work on the analysis of repeated categorical outcome variables provides new methods for the analysis of episode studies. Studies of long-term exposure require comparisons among population groups, and these comparisons can be subject to the design effects characteristic of multistage sample surveys. We examine the implications of these design effects for epidemiologic studies. Finally, the paper discusses the measurement errors induced by the use of outdoor measurements to quantify personal exposure to air pollutants. Recent work on methods for errors-in-variables problems may aid in assessing the effects of such errors on conventional analyses of air-pollution studies.

RÉSUMÉ

Même si les effets écologiques des pluies acides ont été largement signalées, on connaît peu de choses au sujet des effets de la pollution atmosphérique acide sur la santé humaine. Cependant, certaines études épidémiologiques et certaines études impliquant des animaux laissent entendre que l'acidité est un facteur important dans l'explication des troubles respiratoires provoqués par les aérosols. Cet article réexamine quelques aspects de la question et traite de leurs répercussions sur l'élaboration et l'analyse d'études épidémiologiques. Nous mettons en relief deux types de modes d'exposition: les expositions de pointe mais épisodiques à la pollution de l'air et les expositions chroniques à des niveaux constamment élevés de polluants atmosphériques. De nouvelles méthodes d'analyse des études épisodiques sont tirées de travaux récents sur l'analyse des variables répétées de classement en catégories. Les études d'exposition à long terme requièrent des comparaisons entre les groupes et ces comparaisons peuvent être assujetties aux effets du modèle inhérent aux sondages à plusieurs étapes. Nous étudions les conséquences qu'ont ces effets sur les études épidémiologiques.

*Supported by a Cooperative Agreement with the SIAM Institute for Mathematics and Society (SIMS) and the Environmental Protection Agency.

Enfin, cet article traite des erreurs de mesure dues à une prise de mesures à l'extérieur en vue d'évaluer quantitativement une exposition individuelle à la pollution de l'air. Les effets qu'ont de telles erreurs sur les analyses conventionnelles d'études de la pollution atmosphérique peuvent être évalués grâce à des travaux récents sur les problèmes à variables explicatives sujettes à erreur.

1. INTRODUCTION

Although most public discussion of acid rain has focused on the effects of acid precipitation on forests, lakes, and materials, airborne acidity may have adverse health effects as well, especially at the higher concentrations characteristic of air-pollution episodes. This paper begins with a brief review of the evidence for adverse health effects of acid aerosols. Approaches to the design and analysis of epidemiologic studies are then examined. The paper concludes with a discussion of the effects of errors in exposure measurements on epidemiologic studies of air pollutants.

2. HEALTH EFFECTS OF ACIDIC AIR POLLUTION: THE REASONS FOR CONCERN

Although air pollution has been a fact of urban life for centuries, it did not become a major political and public health issue until after World War II. In the early 1950s, The London fogs and oppressive air-pollution episodes in Pittsburgh and other North American cities made it clear that severe air pollution could cause respiratory illness and even threaten life among vulnerable individuals (Ware *et al.*, 1981). These events, along with epidemiologic evidence suggesting adverse health effects of chronic exposure to less severe air pollution, led to programs of air-pollution control in most industrialized countries that have reduced the severity of air pollution in urban areas. One element of those control strategies was the construction of tall smokestacks that produced particulate air pollution of smaller aerodynamic size. These particles can combine with gaseous air pollution to produce acidic particles capable of being transported long distances in the upper atmosphere. In recent years, long-range transported air pollution has been recognized as an important source of ecologic damage to highland lakes and vegetation, and questions are now being raised about its health effects (Franklin *et al.*, 1985)

Lippman (1985) reviewed the evidence for adverse respiratory health effects of ambient acidic aerosols. He argued that, although no direct evidence for toxicity of acid aerosols is available at pollutant concentrations currently seen in North America, strong indirect evidence can be cited for such effects. Schlesinger (1985) found, for example, that single brief exposures to acidic aerosols produced changes in mucociliary clearance rates in animals, while repeated exposures to sulfuric acid (H_2SO_4) produced persistent shifts of clearance rates and changes in the structure of the lung, such as increased numbers of secretory cells and thickened epithelial layers.

No direct epidemiologic evidence is available on the human health effects of acidic air pollution, because technology for field measurement of acidity has only recently become available. One source of indirect evidence is the studies of Lave and Seskin (1977) showing an association between adjusted mortality rates and fossil-fuel air pollutant concentrations in metropolitan areas in the United States. In a recent reanalysis of the original data, Ozkaynak and Spengler (1985) compared four exposure variables: total suspended particulates (TSP), inhalable particles (IP), fine particles (FP), and the sulfate fraction of TSP (SO_4). The first three of these

measurements repre
into the deep lung. T
surrogate for the n
(NH_4HSO_4). Repea
separately with each
found positive coeff
ties

suggesting a stronge
lated with acidity.

Bates and Sizto (1
ments and summert
significant associati
ide (SO_2), ozone (
admissions. The as
lagged by 24 to 48
found that asthma

Finally, Ware *et*
childhood respirat
approximately 10.0
fossil-fuel air pollu
the frequency of r
associated with TS
were found with tw
one second (FEV_1)

Thus, the epider
both from episode
tants. Epidemiolog
analysis, however,
subject to substan
ing different geog
status that may b
discusses some of
research.

3. IMPLICATIONS

Epidemiologic
episodic exposure
elevated concentr
achieve only smal
requirements, as

Studies of acut
course, be self-co
panels of particip
parallels the repe
the association be
such a forced vit

measurements represent successively sharper selection of the small particles inhaled into the deep lung. The fourth, SO_4, also tends to select fine particles and serves as a surrogate for the more acidic sulfates, sulfuric acid and ammonium bisulfate (NH_4HSO_4). Repeating the multiple regression analyses of Lave and Seskin separately with each of these four measures of air pollution, Ozkaynak and Spengler found positive coefficients for all pollutants. The coefficients satisfied the inequalities

$$\beta_{TSP} < \beta_{IP} < \beta_{FP} < \beta_{SO_4},$$

suggesting a stronger relationship as one chooses air pollutants more highly correlated with acidity.

Bates and Sizto (1983) studied the association between hourly pollutant measurements and summertime hospital admissions in southern Ontario. They found highly significant associations between admissions for respiratory disease and sulfur dioxide (SO_2), ozone (O_3), and temperature, but no associations with nonrespiratory admissions. The associations were greatest when the environmental variables were lagged by 24 to 48 hours. In a further analysis of these data, Bates and Sizto (1985) found that asthma admissions had highest correlations with sulfate concentrations.

Finally, Ware *et al.* (1986) reported results from cross-sectional comparisons of childhood respiratory health in six U.S. cities, based on questionnaire data from approximately 10,000 preadolescent children participating in a longitudinal study of fossil-fuel air pollution and respiratory health (Ferris *et al.*, 1979). They found that the frequency of respiratory illnesses and symptoms in the six cities was strongly associated with TSP, SO_4, and sulfur dioxide (SO_2) concentrations. No associations were found with two measures of pulmonary function, forced expiratory volume in one second (FEV_1), and forced vital capacity (FVC).

Thus, the epidemiologic evidence suggests that adverse health effects may result both from episodes and from persistently high concentrations of acidic air pollutants. Epidemiologic studies of air pollutants raise some difficult issues of design and analysis, however, because measures of both exposure and health response are subject to substantial errors of measurement and because ecologic studies comparing different geographic areas are influenced by geographic variability in health status that may be causally unrelated to air-pollution concentrations. This paper discusses some of these issues and strategies for addressing them in air-pollution research.

3. IMPLICATIONS FOR STUDY DESIGN

Epidemiologic studies typically investigate either the acute, reversible effects of episodic exposure to acid aerosols or the chronic effects of long-term exposure to elevated concentrations. Study designs combining both elements are possible, but achieve only small gains in efficiency because of fundamental differences in design requirements, as described below.

Studies of acute effects of fluctuations in acid aerosol concentrations should, of course, be self-controlled. Published studies of air-pollution episodes have enrolled panels of participants and observed them on ten or fewer occasions. This design parallels the repeated-measures analysis of variance, and methods for quantifying the association between occasion-specific values of acidity and *measured* outcomes, such a forced vital capacity, are well known (see Ware 1985, for example). Methods

for the analysis of repeated *categorical* outcomes in this setting are not as well developed, but recent work extends the conceptual and inferential framework of the repeated-measures ANOVA to categorical variables (Liang and Zeger 1986, Stram, Wei, and Ware 1988). This work begins by specifying a mathematical model for the marginal probability of response at each occasion. One familiar choice is the logistic regression model. Although these marginal models cannot be embedded in a natural multivariate distribution for the sequence of outcomes, a "working" correlation matrix for the responses can be specified. Stram, Wei, and Ware use a working assumption of independence, while Liang and Zeger allow an arbitrary working correlation matrix. The estimated regression coefficients derived from maximization of either quasilikelihood can be shown to be asymptotically normal, with asymptotic variances that can be consistently estimated from the data.

Interpretation of studies of pulmonary-function response to air-pollution episodes having very few occasions of measurement has frequently been difficult. When the study begins shortly after an air-pollution episode, for example, there is potential for confounding between air-pollution effects and temporal trends due to learning and adaptation. To avoid the resulting ambiguities, studies of acute responses to air-pollution episodes should begin well before anticipated episodes and require enough repeated measurements to eliminate these temporal trends as an explanatory variable. Episode studies have also tended to be small, primarily because of logistic constraints. Given the relatively small pulmonary-function effects that have been reported, however, future studies should enroll more participants to increase power for detecting these small effects. Carefully designed and executed pulmonary-function examinations are also critically important in these studies.

The acute effects of air-pollution exposure can also be investigated by the collection of daily reports of respiratory symptoms. As part of our longitudinal study, we are collecting year-long records of daily measurements of both air-pollution concentrations and respiratory health status of 1800 participating children. This "diary study" produces a time series of health events for each child. Korn and Whittemore (1979) proposed a two-stage approach to such data. In the initial stage, a regression analysis is performed for each child to produce vectors of child-specific regression coefficients, including a measure of association. At the second stage, these vectors are subjected to further multivariate analysis. Stiratelli, Laird, and Ware (1984) describe a unified logistic-normal mixed model that corresponds to the assumptions implicit in this approach, and propose an algorithm for obtaining approximate maximum-likelihood estimators. Korn and Whittemore's approach is only viable, however, for individuals with sufficient numbers of symptom or illness events to fit the multiple logistic regression model. Moreover, because the number of individual regression coefficients increases as the sample size, both of these methods are computationally burdensome with large data sets.

Alternatively, health responses can be aggregated for each day, and associations between air-pollutant concentrations and changes in symptom status can be investigated using methods for discrete-state stochastic processes. The choice between these two approaches depends in part on the nature of the covariance structure of sequences of symptom states. Our data suggest some heterogeneity of responsiveness among children. When the data on incidence of symptoms among children free of symptoms were analyzed using the model of Williams (1982) for extrabinomial variation, we also found an excess variability of about 10%. Current work is focusing

on modelling of this extrabinomial variability, suggesting some clustering of symptom status within days.

A fundamental design question in diary studies concerns the effects on power of increases in the number of subjects and the number of study days. The answer to that question will depend on this aspect of the covariance structure.

4. ANALYTIC METHODS FOR COMPARISONS AMONG POPULATION GROUPS

Because acid aerosol and fine particle air pollutants are transported over long distances, concentrations are fairly constant over large geographic areas. Thus, studies of the health effects of long-term exposure require comparisons between population groups living in different geographic areas. This creates the potential for "ecological" effects, that is, confounding due to other factors that vary geographically. Ware *et al.* (1986) found that pulmonary-function measurements and symptom states were clustered (positively correlated) within communities. Thus, the unit of observation was the group rather than the individual respondent. This issue corresponds to the "design effect" frequently encountered in sample surveys.

Methods for the analysis of such data are not readily available, especially when large group sizes preclude methods analogous to generalized least-squares regression. The following paragraphs describe one approach to this problem. We begin with the simplest case and then discuss generalizations.

Suppose that Y_{ij} is the response for the jth respondent in the ith group, $j = 1, \ldots, n_i$, $i = 1, \ldots, N$. Associated with individual ij is a vector of covariates, X_{ij}, and a group-specific measure of air-pollution exposure, Z_i. The analysis is performed in two stages. In stage 1, regress Y_{ij} on X_{ij} and indicator variables for groups to produce an adjusted mean, $\hat{Y}_{i\cdot}$, for each group. The first stage may involve either linear or nonlinear regression. Each $\hat{Y}_{i\cdot}$ is assumed to be an unbiased, normally distributed estimator of the true group mean with sampling variance σ_i^2. These variances are estimated at the first stage and then treated as known.

The second-stage model is written as

$$\hat{Y}_{i\cdot} = \beta_0 + \beta_1 Z_i + \gamma_i + \epsilon_i,$$

$i = 1, \ldots, N$. The error terms, γ_i and ϵ_i, are assumed to be independent with variances τ^2 and σ_i^2, respectively. The first of these represents the between-groups variance and is estimated at the second stage.

Under the assumption that the random variables γ_i and ϵ_i, $i = 1, \ldots, N$, are independent and Gaussian, $\hat{Y} = (\hat{Y}_{1\cdot}, \hat{Y}_{2\cdot}, \ldots, \hat{Y}_{N\cdot})$ is distributed as MVN($Z\beta$, Σ), where Z is the design matrix and $\Sigma = G_0 + \tau^2 G_1$. G_0 is the $M \times M$ diagonal matrix of variances, σ_i^2, and G_1 is the $M \times M$ identity matrix. Differentiating the log likelihood with respect to τ^2, we obtain the estimating equation

$$\mathrm{tr}(\Sigma^{-1} G_1) - \mathrm{tr}(\Sigma^{-1} G_1 \Sigma^{-1} C) = 0,$$

where $C = (Y - Z\beta)(Y - Z\beta)'$. Differentiating with respect to β and taking expectations shows that the estimate of β is asymptotically uncorrelated with that of τ^2, while differentiating with respect to τ^2 again and taking expectations gives the expected information matrix for the estimate of τ^2.

Fisher's scoring procedure starts with an initial estimate of τ^2, then alternates between recalculation of $\hat{\beta} = (Z'\hat{\Sigma}^{-1}Z)^{-1}Z'\hat{\Sigma}^{-1}Y$ and $C = (Y - Z\hat{\beta})(Y - Z\hat{\beta})'$ and

reestimation of τ^2 by

$$\hat{\tau}^{2(l+1)} = \hat{\tau}^{2(l)} - \{\mathrm{tr}(\hat{\Sigma}^{-1}G_1\hat{\Sigma}^{-1}G_1)\}^{-1}(\mathrm{tr}(\hat{\Sigma}^{-1}G_1) - \mathrm{tr}(\hat{\Sigma}^{-1}G_1\hat{\Sigma}^{-1}C))$$

with $\hat{\Sigma} = \hat{\Sigma}^{(l)}$. When convergence is achieved, inference about β is based on the asymptotic covariance matrix $(Z'\hat{\Sigma}^{-1}Z)^{-1}$, while inference concerning τ^2 is based on $0.5(\mathrm{tr}(\hat{\Sigma}^{-1}G_1\hat{\Sigma}^{-1}G_1))^{-1}$.

Notice that $\Sigma = G_0 + \tau^2 G_1$ is linear in τ^2, but includes G_0 as a fixed known contribution to the variances. The estimation procedure outlined above may be extended so that Σ takes the form

$$G_0 + \sum_{i=1}^{q} \tau_i^2 G_i, \tag{5.1}$$

where again the matrices G_i, $i = 0, \ldots, q$, are known and τ_i^2, $i = 1, \ldots, q$, are to be estimated. The linear pattern of the second term in (5.1) allows for a wide variety of additional random effects. For example, there may be several distinct subgroups to be nested within each geographic group, but the clustering may remain at the group (i.e. geographic) level rather than the subgroup. In this case, the corresponding G matrix would be block diagonal with blocks of 1's modelling the nesting of subgroups within geographic group. Ware *et al.* (1986) used covariance matrices like (5.1) to allow nesting of age cohort within city and to investigate the possibility that there were separate (random) within-city regression coefficients of health outcome against air pollution for each city. The corresponding G matrix consisted of blocks along the diagonal with the ith block being the sum-of-squares-and-cross-products matrix $(z_{ij} - z_{i\cdot})(z_{ik} - z_{i\cdot})$, $j, k = 1, \ldots, n_i$, with n_i the number of cohorts within city i and z_{ik} the air-pollution measurement in city i for the kth cohort.

An important generic issue involved in studies of this kind is the determination of the population of interest and the proper experimental unit. Ware *et al.* (1986) treated the cities as a sample of size 6 from the population of all U.S. cities. A crucial issue in study design is the tradeoff between the numbers of subjects to be examined within each location and the total number of locations to be sampled. Sample-size calculations require information about the relative sizes of the variance components.

To examine the effect of clustering of illness response within cities, we considered a hypothetical design involving samples of size 200 or 400 in each of five or ten cities nested within five regions. We assumed that the dose-response relationship was logistic-linear in air-pollution concentration and that the frequency of the endpoint under study, persistent cough, was 0.06 at the minimum air-pollution exposure. The clustering effect was assumed to arise at the city level, and regions were treated as fixed effects. The strength of the association between air-pollution concentration and frequency of cough was expressed as the relative odds for cough at the upper and lower ends of the range of exposures. Assuming a uniform distribution of air-pollution concentrations over the five regions, we examined the power of a study with different combinations of sample size within towns, number of towns, between-city variance component, and effect (Table 1).

Two values of the between-city variance, 0 and 0.04, were considered, the latter chosen to correspond to the value found in an analysis of data from the Six Cities Study. Samples of size 400 in each city in the presence of a random effect give power comparable to samples of size 200 in the absence of such an effect. Ware *et al.* found that the between-city variance components were large, resulting in substantial loss of

TABLE 1: P
frequency
variance: n
upper and

σ^2
0
0.04

[a]The base
depends
formed at

precision relative to
tion of such cluster
study design. Betwe
number of study co

5. THE EFFECTS
EXPOSURE

Despite an exte
analysis, errors in
when they are ve
because measurer
massive errors wh
progress on the u
personal exposure
issue.

The generic is
statistical relation
but a third varia
formulation, the
where E_i is a sequ
usually centered
with measureme
parameters β by u
"naive" estimate

In air-pollutio
ments are less pr
toring in individ
from personal ex
existence and rel
toring, it is appar
This perspective

TABLE 1: Power calculations for effect of exposure to air pollutants on frequency of cough.[a] K = number of cities per region; σ^2 = between-cities variance; n = number of children per city; OR = relative odds of cough at the upper and lower limits of the range of air pollutant concentrations.

		Power			
		$K = 5$		$K = 10$	
σ^2	OR	$n = 200$	$n = 400$	$n = 200$	$n = 400$
0	1.2	0.22	0.38	0.38	0.65
	1.4	0.64	0.90	0.90	0.99
	1.6	0.93	0.99	0.99	0.99
0.04	1.2	0.16	0.21	0.27	0.37
	1.4	0.45	0.59	0.73	0.87
	1.6	0.75	0.88	0.96	0.99

[a]The baseline frequency of cough is assumed to equal 0.06. The power depends on air-pollutant concentrations only through OR. The test is performed at the two-sided 0.05 level.

precision relative to the naive analysis. These results point to the need for identification of such clustering as it occurs in health data, and an appropriate response in the study design. Between-city variability, for example, implies the need to increase the number of study communities.

5. THE EFFECTS OF ERRORS IN MEASUREMENT OF AIR-POLLUTION EXPOSURE

Despite an extensive literature on errors in independent variables in regression analysis, errors in independent variables tend to be ignored in applied work, even when they are very large. This is especially surprising in air-pollution research, because measurements of air-pollution concentrations at outdoor stations have massive errors when treated as estimates of personal air-pollution exposure. Recent progress on the use of personal and indoor monitoring to improve estimates of personal exposure (Leaderer *et al.* 1986) has, however, created new interest in this issue.

The generic issue in errors-in-variables problems is that of establishing the statistical relationship between two variables, say Y and X, when X is not observed but a third variable, Z, stochastically related to X, is observed. In the classical formulation, the relationship between X and Z is additive, that is, $Z_i = X_i + E_i$, where E_i is a sequence of independent and identically distributed random variables, usually centered at 0. Here, X is the "true" convariate, and Z is X contaminated with measurement error. If the model for Y is $Y_i = f(X_i|\boldsymbol{\beta}) + e_i$, then estimating the parameters $\boldsymbol{\beta}$ by using Z in place of X will lead to biased estimates, usually called the "naive" estimates. It is usual to assume that E_i is independent of e_i.

In air-pollution epidemiology, exposure estimates based on ambient measurements are less precise than estimates incorporating information from indoor monitoring in individual homes, and these are less precise than measurements obtained from personal exposure monitors worn throughout the day. Once one recognizes the existence and relevance of exposure estimates based on continuous personal monitoring, it is apparent that the measurements available for analysis are subject to error. This perspective raises doubts about the assumption of additive errors, because there

will, in general, be systematic as well as random errors involved in using ambient measurements or indoor monitoring to estimate personal air-pollution exposures. The situation can be represented more generally by the notion of a *surrogate variable* rather than an errors-in-variables problem. Surrogate variables may be subject to systematic as well as zero-centered random errors. Z is a surrogate for X if the outcome, Y, is conditionally independent of Z, given X.

If the primary goal in a regression analysis is testing of the null hypothesis of no association between air pollution and health, errors in exposure measurements are of concern only because of their effect on sensitivity. This is true because an important class of hypothesis tests, namely score tests, will, under broad conditions, have the intended Type I error rate asymptotically when Z is used in place of X in all calculations (Tosteson and Tsiatis 1988). This point has not always been properly understood in the literature. For example, Pickles (1982) claimed that, in a univariate linear regression analysis of the relationship between SO_2 concentration and mortality in 16 British towns, random errors in the SO_2 measurements implied that the P-value for the hypothesis of no association was biased toward 0. While the naive regression coefficient, $\hat{\beta}$, was biased towards 0 in his example, the variance of the naive estimator was consistently estimated by standard techniques. Thus, measurement errors decreased the power of the test but did not affect the null distribution of the test statistic.

Because these results on the nominal size of the naive score tests hold only asymptotically, small-sample tests relying on the parametric form of the conditional distribution of Y given X may be inappropriate if a surrogate variable is measured instead of X. In general, the error structure of the conditional distribution of Y given Z will not be of the same parametric family as that for Y given X. This may have important implications for inference in small samples. As a practical matter, however, an epidemiologist is seldom in the happy position of knowing exactly the form of the error distribution even without the additional complications of errors in variables. Diagnostic checking for lack of fit of regression-model assumptions is always advisable whether or not the independent variable is observed with error.

When one considers more general analytic objectives than the testing of null hypotheses in univariate regression models, errors in variables and surrogate variables raise more difficult issues. The size of the naive score test remains correct in large samples when testing the null hypothesis that all coefficients in a multiple regression model equal 0. The size of tests of "lower-dimensional" null hypotheses may be disturbed by bias away from the null in individual parameter estimates.

When the objective is estimation, a number of procedures can be utilized to assess the bias caused by errors in variables. Direct attempts to reduce the bias of naive estimators can raise difficult computational problems even in situations where the joint distribution of measurement errors and the covariate X is completely known. Maximum likelihood in such cases is often extremely difficult, with numerical integration required over high-dimensional regions. Methods for approximating the maximum-likelihood estimate with much less calculation are discussed by Stefanski (1985) and Whittemore and Keller (1986). These methods require knowledge about the distribution of the measurement error which is not always available. The methods are most appropriate for gaining insight about the effects of measurement errors over a range of possible measurement error probability models, rather than as a device for obtaining a single "right number" to serve as a revised estimate. The availability of these new methods for quantifying the effects of errors in variables on

regression anal
exposure estim
opportunity fo
errors in variab

REFERENCES

Bates, D.V., and S
southern Ont
Bates, D.V., and S
Aerosols: Res
P.J. Verkert, e
Ferris B.G., Jr.; Sp
on human he
767.
Franklin, C.; Bur
Canadian pe
Korn E.L., and W
pollution. *Bi*
Lave, L.B., and S
Leaderer, B.P.; Z
indoor air co
124, 275–28
Liang, K.-Y., an
Biometrika,
Lippman, M. (198
63, 63–70.
Ozkaynak, H., an
acid precipit
Pickles, J.H. (198
Env., 16, 224
Schlesinger, R.B.
Persp., 63, 2
Stefanski, L.A. (
583–592.
Stiratelli, R.; Lai
binary respo
Stram, D.O.; We
possibly mis
Tosteson, T.D., "
Biometrika
Ware, J.H. (198
Ware, J.H.; Ferri
ambient sul
Rev. Resp. I
Ware, J.H.; Thi
health effec
studies. *Em*
Whittemore, A.
lished.
Williams, D.A.

Received 20 No
Revised 15 July
Accepted 20 Jul

regression analyses, and the increasing understanding of the magnitude of errors in exposure estimation associated with different measurement strategies, creates an opportunity for empirical research aimed at improving our ability to account for errors in variables in the analysis of epidemiologic data.

REFERENCES

Bates, D.V., and Sizto, R. (1983). Relationship between air pollutant levels and hospital admissions in southern Ontario. *Canad. J. Pub. Hlth.*, 74, 117–122.

Bates, D.V., and Sizto, R. (1986). A study of hospital admissions and air pollutants in southern Ontario. *Aerosols: Research, Risk Assessment and Control Strategies* (S. D. Lee, T. Schneider, L.D. Grant, and P.J. Verkert, eds.), Lewis Publishers, Chelsea Michigan, 767–777.

Ferris B.G., Jr.; Speizer, F.E.; Spengler, J.D., *et al.* (1979). Effects of sulfur oxides and respirable particles on human health: Methodology and demography of populations in study. *Amer. Rev. Resp. Dis.*, 120, 767.

Franklin, C.; Burnett, R.T.; Paolini, R.J.P., and Raizenne, M.E. (1985). Health risk from acid rain: A Canadian perspective. *Env. Hlth. Persp.*, 63, 155–168.

Korn E.L., and Whittemore, A.S. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics*, 35, 795–802.

Lave, L.B., and Seskin, E.P. (1977). *Air Pollution and Human Health*. Johns Hopkins Press, Baltimore.

Leaderer, B.P.; Zagraniski, R.T.; Berwick, M., and Stolwijk, J.A.J. (1986). Assessment of exposure to indoor air contaminants from combustion sources: Methodology and application. *Amer. J. Epid.*, 124, 275–289.

Liang, K.-Y., and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.

Lippman, M. (1985). Airborne acidity: Estimates of exposure and human health effects. *Env. Hlth. Persp.*, 63, 63–70.

Ozkaynak, H., and Spengler, J.D. (1985). Analysis of health effects resulting from population exposures to acid precipitation precursors. *Env. Hlth. Persp.*, 63, 45–55.

Pickles, J.H. (1982). Air pollution estimation error and what it does to epidemiological analysis. *Atmos. Env.*, 16, 2241–2245.

Schlesinger, R.B. (1985). Effects of inhaled acids on respiratory tract defense mechanisms. *Env. Hlth. Persp.*, 63, 25–38.

Stefanski, L.A. (1985). The effects of measurement error on parameter estimation. *Biometrika*, 72, 583–592.

Stiratelli, R.; Laird, N.M., and Ware, J.H. (1984). Random effects models for serial observations with binary response. *Biometrics*, 40, 961–972.

Stram, D.O.; Wei, L.J., and Ware, J.H. (1988). Analysis of repeated ordered categorical outcomes with possibly missing observations. *J. Amer. Statist. Assoc.*, to appear.

Tosteson, T.D., Tsiatis, A.A. (1986). The use of a surrogate independent variable in tests for association. *Biometrika*, to appear.

Ware, J.H. (1985). Linear models for the analysis of longitudinal studies. *Amer. Statist.*, 39, 95–101.

Ware, J.H.; Ferris, B.G. Jr.; Dockery, D.W.; Spengler, J.D.; Stram, D.O., and Speizer, F.E. (1986). Effects of ambient sulfur oxides and suspended particles on respiratory health of preadolescent children. *Am. Rev. Resp. Dis.*, 133, 834–842.

Ware, J.H.; Thibodeau, L.A.; Speizer, F.E.; Colome, S., and Ferris, B.G., Jr. (1981). Assessment of the health effects of atmospheric sulfur oxides and particulate matter: Evidence from observational studies. *Env. Hlth. Persp.*, 41, 255–276.

Whittemore, A.S., and Keller, J.B. (1986). Approximations for errors-in-variables regression. Unpublished.

Williams, D.A. (1982). Extra-binomial variation in logistic-linear models. *Appl. Statist.*, 31, 144–148.

Received 20 November 1986
Revised 15 July 1987
Accepted 20 July 1987

Department of Biostatistics
School of Public Health
Harvard University
Boston, Massachusetts 02115
U.S.A.

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This article documents research that was conducted under the auspices of the U.S. Environmental Protection Agency (USEPA). As an arm of the government, the USEPA has a paid up, non-exclusive, irrevocable, worldwide license to reproduce, prepare derivative works, and distribute to the public all of the material contained herein, and to grant others acting on its behalf the right to do so. The USEPA grants to the National Technical Information Service, Department of Commerce, the right to act as its agent in the distribution of this material.